UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/814,382	03/31/2004	Gautam Das	307517.01	5509

22971   7590   05/03/2007
MICROSOFT CORPORATION
ONE MICROSOFT WAY
REDMOND, WA 98052-6399

EXAMINER
LOVEL, KIMBERLY M

ART UNIT	PAPER NUMBER
2167	

NOTIFICATION DATE	DELIVERY MODE
05/03/2007	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

roks@microsoft.com
ntovar@microsoft.com
a-rydore@microsoft.com

PTOL-90A (Rev. 04/07)

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>31 March 2004</u>.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) <u>1-34</u> is/are pending in the application.

    4a) Of the above claim(s) <u>29-31,33 and 34</u> is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☐ Claim(s) _____ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on <u>31 March 2004</u> is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO/SB/08)
Paper No(s)/Mail Date <u>8/19/05</u>.

4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

1.   Claims 1-28 and 32 are rejected.

Election/Restrictions

2.   Applicant's election of Group I (claims 1-28 and 32) in the reply filed on 20

February 2007 is acknowledged. Because applicant did not distinctly and specifically

point out the supposed errors in the restriction requirement, the election has been

treated as an election without traverse (MPEP § 818.03(a)).

Information Disclosure Statement

3.   The information disclosure statement (IDS) submitted on 19 August 2005 was

filed after the mailing date of the application on 31 March 2003. The submission is in

compliance with the provisions of 37 CFR 1.97. Accordingly, the information disclosure

statement is being considered by the examiner.

/

Claim Rejections - 35 USC § 102

The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that

form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

4.   **Claims 1-28 and 32 are rejected under 35 U.S.C. 102(e) as being anticipated**

by US Pat No 6,865,567 to Oommen et al (hereafter Oommen).

Referring to claim 1, Oommen discloses in a database system, a sampling

method for constructing a data structure based on the contents of a database

comprising:

a) gathering an initial sample [first phase] of data from the database and creating

a first data structure from said initial sample [x number of tuples] (see column 21, lines

16-19);

b) gathering a second sample [second phase] of data from the database (see

column 21, lines 19-28);

c) determining an initial sufficiency of the data gathered from the database that is

based on a comparison of the first data structure and the second sample of data (see

column 21, lines 19-40); and

d) forming a resultant data structure by gathering an additional sample of data

from the database and using the additional amount of data to form the resultant data

structure wherein the amount of data gathered in the additional sample is based on the

initial sufficiency determination (see column 22, line 48 – column 23, line 17).

Referring to claim 2, Oommen discloses the method of claim 1 wherein the

resultant data structure is formed based on data gathered in the initial sample, the

second sample and the additional sample (see column 21, lines 19-40 and column 22,

line 48 – column 23, line 17).

Referring to claim 3, Oommen discloses the method of claim 1 wherein the first

and resultant data structures are histograms (see column 22, lines 60-67).

Referring to claim 4, Oommen discloses the method of claim 1 wherein the

initial and second data samples are randomly retrieved block samples that form a first

amount of data that is initially gathered and then divided in half to provide the initial and

second data samples (see column 20, lines 60-67).

Referring to claim 5, Oommen discloses the method of claim 4 wherein the

initial and second data samples are sorted and used to form two histograms (see Fig

11).

Referring to claim 6, Oommen discloses the method of claim 5 wherein an error

metric of the two histograms are formed by cross correlating the contents of the two

histograms to determine the initial sufficiency (see Fig 19).

Referring to claim 7, Oommen discloses the method of claim 6 wherein the

initial and second data samples are further sub-divided to form sub-samples used to

form other histograms of differing sample sizes that are cross correlated to find an error

metric relating to said differing sample sizes (see Fig 19).

Referring to claim 8, Oommen discloses the method of claim 6 wherein the initial and second data samples are further sub-divided to form additional sub-samples of smaller size that are used to form other histograms that are cross correlated for use in finding an error metric relating to sample sizes for use in determining a size of the additional sample of data to gather from the database (see Fig 11).

Referring to claim 9, Oommen discloses the method of claim 4 additionally comprising estimating distinct values of an attribute of the initial and second samples by eliminating records from the blocks that are duplicated within a given block and estimating distinct values by categorizing attributes as rarely or frequently occurring within the database (see column 7, lines 40-49).

Referring to claim 10, Oommen discloses a computer readable medium for performing computer instructions to implement the method of claim 1 (see column 117, lines 12-24).

Referring to claim 11, Oommen discloses a database system for constructing histograms based on sampling the contents of the database comprising:

a) a database management component that gathers block size data segments from the database which in aggregate form a first sample of data having a first size [first phase – x number of tuples] (see column 21, lines 16-19);

b) a histogram construction component that forms a first histogram from the first sample of data (see Fig 11); and

c) a correlation component that determines an initial sufficiency of the first

sample of data gathered from the database based on a comparison of the first

histogram and data from the first sample of data (see column 21, lines 19-40);

d) wherein said database management component gathers an additional sample

of data used by said histogram construction component in creating a resultant

histogram and the size of the additional sample is based on the initial sufficiency

determination (see column 22, line 48 – column 23, line 17).

Referring to claim 12, Oommen discloses the system of claim 11 wherein the

resultant histogram is formed by the histogram construction component based on data

gathered in the first sample of data and the additional data (see column 22, lines 60-67).

Referring to claim 13, Oommen discloses the system of claim 11 wherein the

first sample of data and the additional sample of data are randomly retrieved block

samples (see column 20, lines 60-67).

Referring to claim 14, Oommen discloses the system of claim 11 wherein

histogram construction component sorts the data in said first sample of data as it

constructs the first histogram (see column 22, lines 60-67 and Fig 11).

Referring to claim 15, Oommen discloses the system of claim 11 wherein the

correlation component determines an error metric by cross correlating the contents of

the first histogram with other data in said first sample of data to determine the initial

sufficiency (see Fig 11 and Fig 19).

Referring to claim 16, Oommen discloses the system of claim 15 wherein the

first sample of data is sub-divided to form sub-samples used to form histograms of

differing sizes that are cross correlated to find an error metric relating to said differing

sample sizes (see Fig 19).

Referring to claim 17, Oommen discloses the system of claim 15 wherein the

first sample of data is sub-divided to form additional sub-samples of smaller size that

are used to form other histograms that are cross correlated for use in finding an error

metric relating to sample sizes for use in determining a size of the additional sample of

data to gather from the database (see Fig 11).

Referring to claim 18, Oommen discloses in a database system, a sampling

method for constructing a histogram based on the contents of a database comprising:

a) gathering an initial sample [first phase with x number of tuples] (see column

21, lines 16-19 and column 22, lines 60-67) of data from the database and creating a

histogram from said initial sample;

b) gathering a second sample of data from the database for comparison with said

first histogram [second phase] (see column 21, lines 19-40);

c) determining an initial sufficiency of the data gathered from the database that is

based on a comparison of the second sample with the first histogram (see column 21,

lines 19-40); and

d) if the determination of initial sufficiency indicates the data in said initial and

second samples is adequate to represent the database, combining the initial and

second samples to form a resultant histogram, but if the determination of initial

sufficiency indicates the initial and second samples are inadequate to represent the

database, gathering an additional data sample to combine with the initial and second

samples to form the resultant histogram wherein a size of the additional data sample is based on the initial sufficiency determination (see column 22, line 48 – column 23, line 17).

Referring to claim 19, Oommen discloses the method of claim 18 wherein the data is gathered in blocks from random storage locations within the database (see column 20, lines 60-67).

Referring to claim 20, Oommen discloses in a database system, a system for constructing a data structure based on the contents of a database comprising:

a) means for gathering an initial sample [first phase] of data from the database and creating a first data structure [histogram] from said initial sample (see column 21, lines 16-19 and column 22, lines 60-67);

b) means for determining an initial sufficiency of the data gathered from the database that is based on a comparison of the first data structure and other data in the initial sample not used to create the first data structure (see column 21, lines 19-40); and

c) means for forming a resultant data structure by gathering an additional sample of data from the database and using the additional amount of data to form the resultant data structure wherein the amount of data gathered in the additional sample is based on the initial sufficiency determination (see column 22, line 48 – column 23, line 17).

Referring to claim 21, Oommen discloses the system of claim 20 wherein the resultant data structure is formed based on data gathered in the initial sample and the additional sample (see column 21, lines 19-40 and column 23, line 17).

Referring to claim 22, Oommen discloses the system of claim 21 wherein the first and resultant data structures are histograms (see column 22, lines 60-67).

Referring to claim 23, Oommen discloses the system of claim 20 wherein the initial data sample is made up of randomly retrieved block samples that form a first amount of data that is divided in half to provide data to form the data structure and data to cross correlate against the first data structure (see column 20, lines 60-67).

Referring to claim 24, Oommen discloses the system of claim 23 wherein the initial data samples is sorted and used to form two histograms (see Fig 11).

Referring to claim 25, Oommen discloses the system of claim 24 wherein an error metric of the two histograms are formed by cross correlating the contents of the two histograms to determine the initial sufficiency (see Fig 19).

Referring to claim 26, Oommen discloses the system of claim 25 wherein the initial data sample is further sub-divided to form sub-samples used to form other histograms of differing sample sizes that are cross correlated to find an error metric relating to said differing sample sizes (see Fig 19).

Referring to claim 27, Oommen discloses the system of claim 26 wherein the initial and second data samples are further sub-divided to form additional sub-samples of smaller size that are used to form other histograms that are cross correlated for use in finding an error metric relating to sample sizes for use in determining a size of the additional sample of data to gather from the database (see Fig 11).

Referring to claim 28, Oommen discloses the system of claim 24 additionally comprising means for estimating distinct values of an attribute of the initial and second
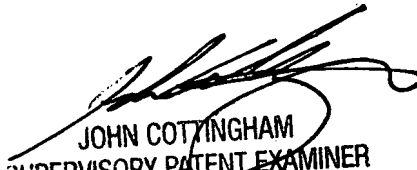
samples by eliminating records from the blocks that are duplicated within a given block and estimating distinct values by categorizing attributes as rarely or frequently occurring within the database (see column 7, lines 40-49).

Referring to claim 29, Oommen discloses a computer readable medium for performing computer instructions to implement the method of claim 20 (see column 117, lines 12-24).

Conclusion

5.   The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

- US Patent No 6,785,684 titled "Apparatus and Method for Determining Clustering Factor in a Database using Block Level Sampling"

JOHN COTTINGHAM
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100

Contact Information

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Kimberly Lovel whose telephone number is (571) 272-2750. The examiner can normally be reached on 8:00 - 4:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Cottingham can be reached on (571) 272-7079. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

Kimberly Lovel
Examiner
Art Unit 2167

26 April 2007
kml